Thanks for Dangna Li for providing these solutions.

Problem 1

1. By Neyman-Pearson factorization criterion, T is sufficient iff there exists a function $g_P(\cdot)$ such that

$$\frac{\prod_{i=1}^n f_P(x_i)}{\prod_{i=1}^n f_{P_0}(x_i)} = g_P(T(x^{(n)}))$$

Taking log on both sides yields:

$$\varphi(x_1) + \cdots + \varphi(x_n) = \log g_P(T(x^{(n)})) \triangleq \psi(T(x^{(n)}))$$

i.e. $\varphi_p \in \mathcal{S}$. In other words, T is sufficient if and only if $\varphi_p \in \mathcal{S}$.

2. Since by definition $\mathbb{P}$ is exponential of order s if and only if s is the smallest positive integer such that

$$f_P(x) = exp\left\{\sum_{i=1}^s \eta_i(P)\tau_i(x) - B(P)\right\} h(x)$$

then it is easy to prove that $dim\{span\{\tau_1, \cdots, \tau_s\}\} = s$, where $span\{\tau_1, \cdots, \tau_s\}$ is the linear space spanned by $\tau_1, \cdots, \tau_s$. This is because if $dim\{span\{\tau_1, \cdots, \tau_s\}\} < s$, then there exists an integer $s' < s$, such that $f_P(x)$ admits the above form.

On the other hand, notice

$$\varphi_p(x) = B(P_0) - B(P) + \sum_{i=1}^S (\eta_i(P) - \eta_i(P_0))\tau_i(x)$$

Thus we have $V = span\{\mathbb{I}, \tau_1, \cdots, \tau_s\}$, where $\mathbb{I}$ is the constant function. [*ed. note:* The above shows $V \subseteq span\{\mathbb{I}, \tau_1, \cdots, \tau_s\}$, and V cannot be of lower dimension, or else we could work backwards and write the densities $f_P(x)$ using fewer τ_i.] To sum up, $dim(V) = s + 1$ if and only if P is of order s.

3. Let $t_1 = T(x_0, y_1), t_2 = T(x_0, y_2)$. Without loss of generality, we shall assume $y_1 < y_2$ and

$$t_1 < T(x_0, y) < t_2, \quad y \in (y_1, y_2) \tag{1}$$

Otherwise, we can always define $y_1' = \sup\{y | y \leq y_2, T(x_0, y) = t_1\}$, $y_2' = \inf\{y | y \geq y_1, T(x_0, y) = t_2\}$. Then by redefining $y_1 = y_1', y_2 = y_2'$, we have $T(x_0, y_1) = t_1$, $T(x_0, y_2) = t_2$, and $\varphi(y_1) = \varphi(y_2)$.

Since $\varphi(y_1) = \varphi(y_2)$ and φ is continuous, then there exists a $y_0 \in (y_1, y_2)$ such that y_0 is a local maximal or local minimum. Without loss of generality, suppose

$$\varphi(y_0) \geq \varphi(y), \quad y \in [y_1, y_2] \tag{2}$$

The other case can be proved analogously.

From (1), we knot that

$$t_1 < t_0 = T(x_0, y_0) < t_2$$

Moreover,

$$\psi(t_0) = \varphi(y_0) + \varphi(x_0) \geq \varphi(y) + \varphi(x_0)$$

for all $y \in (y_1, y_2)$, where the last inequality follows from (2)

Therefore it follows from (1) that

$$\psi(t_0) \geq \psi(t), \quad t \in (t_1, t_2) \tag{3}$$

Thus there exists an interval $U(x_0, \epsilon)$, which is of radius ϵ and around x_0, such that for all $x \in U(x_0, \epsilon)$, we have

$$T(x, y_1) < t_0 < T(x, y_2) \tag{4}$$

and

$$t_1 < T(x, y_0) < t_2 \tag{5}$$

By continuity of T and (4), we know that $\forall$ x, there exists a $y_x \in (y_1, y_2)$, such that $T(x, y_x) = t_0$

Hence when $x \in U(x_0, \epsilon)$, we have

$$\begin{aligned} \psi(t_0) = \psi(T(x, y_x)) &= \varphi(x) + \varphi(y_x) && (6)\\ &\leq \varphi(x) + \varphi(y_0) = \psi(T(x, y_0)) && (7)\\ &\leq \psi(t_0) = \varphi(x_0) + \varphi(y_0) && (8) \end{aligned}$$

where the inequality in (7) follows from (2) and the inequality in (8) follows from (3) and (5). Moreover, none of the inequality above can hold strictly. Thus

$$\varphi(x) = \varphi(x_0), \quad x \in U(x_0, \epsilon)$$

4. Take any φ_1 and $\varphi_2 \in \mathcal{S}$, we show that there exists a constant a such that $\varphi = \varphi_1 - a\varphi_2$ and φ satisfies the conditions in previous part. Therefore by the result from that part we know that $\varphi(x) = \varphi(x_0)$ when x is in some small neighborhood of x_0 for any fixed x_0. Since x_0 is arbitrary, by continuity of ψ, we can conclude there must exist a constant b such that $\varphi = b$. i.e. $\varphi_2 = a\varphi_1 + b$. This implies $dim(V) \leq dim(S) \leq 2$, which further implies $\mathbb{P}$ is of order 1.

   We now prove the existence of φ:

   proof: Without loss of generality, suppose φ_1 is not a constant function. Then there must exists $y_1 \neq y_2$ such that $\varphi_1(y_1) \neq \varphi_1(y_2)$.

   Therefore
$$T(x_0, y_1) \neq T(x_0, y_2) \quad \forall x_o \in \mathcal{X}$$

   Moreover, if we let $a = \frac{\varphi_2(y_1)-\varphi_2(y_2)}{\varphi_1(y_1)-\varphi_1(y_2)}$, then it follows that

$$\varphi_2(y_1) - a\varphi_1(y_1) = \varphi_2(y_2) - a\varphi_1(y_2)$$

   Now let $\varphi = \varphi_2 - a\varphi_1$. By the discussion at the beginning of the proof, we have $\varphi = b$ for some constant b. Since φ_1, φ_2 are arbitrarily chosen, one can conclude that $dim(S) \leq 2$

Problem 2

1. The log likelihood function is given by

$$l(\theta) = \sum_i -\frac{1}{2}\frac{(y_i - \theta)^2}{\sigma_{a_i}^2}$$

   Taking derivative and set it to zero, we have

$$\sum_i \frac{y_i}{\sigma_{a_i}^2} = \hat{\theta}\sum_i \frac{1}{\sigma_{a_i}^2}$$

   Therefore

$$\hat{\theta} = \frac{\sum_i y_i/\sigma_{a_i}^2}{\sum_i 1/\sigma_{a_i}^2}$$

2. Since $l''(\theta) = -1/\sigma_a^2$, hence

$$I(\theta) = n\mathbb{E}_{(y,a)}(1/\sigma_a^2) = n\left(\frac{p}{\sigma_1^2} + \frac{1-p}{\sigma_2^2}\right)$$

3. By asymptotic theory of MLE we know that $\hat{\theta}$ is asymptotically distributed as $\mathcal{N}(\theta, 1/I(\theta))$. Since in out setting $I(\theta)$ does not depend on θ and hence can be used for inference. However, in this case it is obvious that the conditional variance of $\hat{\theta}$ given $(a_1, \cdots, a_n)$, which is given by the inverse of the observed fisher information, a.k.a $1/\hat{I}(\theta) = (\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2})^{-1}$, is a more meaningful quantity.

   If we use the unconditional variance to form a confidence interval, it would not have the right coverage. For example, let $\sigma_1^2 = 100$, $\sigma_2^2 = 1$, $n_1 = 9, n_2 = 1$, $p = 0.6$ then if we use the unconditional variance, the approximate 95% confidence interval is given by

$$\hat{\theta} \pm 1.96/\sqrt{I(\theta)} = \hat{\theta} \pm 1.96/\sqrt{5.05}$$

   where as the correct 95% confidence interval (based on conditional variance) is

$$\hat{\theta} \pm 1.96/\sqrt{\hat{I}(\theta)} = \hat{\theta} \pm 1.96/\sqrt{1.09}$$

   Admittedly this example is quite extreme in that the experimenter was really unlucky in picking the bad equipment 9 times out of 10. But the main point is that as statisticians, we should always use the relevant variance estimate.